These estimates for each NCTR pathologist are then averaged to obtain an overall standardized estimate at dose d for the SOT pathologist.

References

1. CAIRNS T. The ED_{01} study: introduction, objectives, and experimental design. J Environ Pathol Toxicol 1980;3:1–7.
2. Society of Toxicology. Re-examination of the ED_{01} study audit of pathology. Fundam Appl Toxicol 1981;1:64–66.
3. COX DR. Analysis of binary data. London: Methuen, 1970.
4. LAGAKOS SW, RYAN LM. On the representativeness assumption in prevalence tests of carcinogenicity. J R Stat Soc C 1985;34:54–62.
5. National Toxicology Program. Report on the carcinogenesis bioassay of Chrysotile asbestos in Syrian golden hamsters, National Toxicology Program report, 1981 [DHHS publication No. (TR)81-246].
6. WARD JM. Background data and variations in tumor rates of control rats and mice. In: Homburger F, ed. Progress in experimental tumor research. Basel: Karger, 1983:241–258.
7. PARKER CM, VAN GELDER GA, CHAI EY, et al. Oncogenic evaluation of tetrachlorvinphos in the $B6C3F_1$ mouse. Fundam Appl Toxicol 1985;5:840–854.
8. STANLEY KE, MATTHEWS MJ. Analysis of a pathology review of patients with lung tumors. JNCI 1981;66:989–992.
9. WARD JM, GOODMAN DG, GRIESEMER RA, et al. Quality assurance for pathology in rodent carcinogenesis tests. J Environ Pathol Toxicol 1978;2:371–378.

Examination of the Role of Cigarette Smoke in Lung Carcinogenesis Using Multistage Models[1,2]

Michael Gaffney,[3] *Bernard Altshuler*[4,5]

The widely used multistage model of Armitage and Doll is fit to the British physician lung cancer data of Doll and Hill under the assumption that cigarette smoke induces the initial and penultimate changes. It is shown that the best fit of this model in continuing smokers gives predictions not in accordance with incidence in ex-smokers and dose-response. A better global fit can be obtained by increasing the number of stages, but this de-emphasizes initiation and is inconsistent with the rise of incidence in nonsmokers. Thus, one should look to other models. A two-stage model with clonal growth in which smoking initiates normal target cells and promotes the clonal growth of just the smoke-initiated cells is proposed. This model is shown to agree with the Doll and Hill data and thus it has empirical plausibility that should encourage biological studies of clonal growth in carcinogenesis. [J Natl Cancer Inst 1988; 80:925–931]

The best data relating exposure to cancer in humans is that obtained by Doll and Hill on cigarette smoking and lung cancer in British physicians (*1,2*). These data have been used to examine the effects of smoking duration, amount smoked, and stopping smoking on the incidence of lung cancer. As a result of these analyses, the etiological role that smoking plays in lung carcinogenesis has been interpreted in the context of the multistage model of Armitage and Doll (*3*). The essence of this model is that an ordered sequence of discrete cellular changes is needed to transform a normal cell into a malignant one without proliferation of intermediate stages.

As a result of the observed time patterns of lung cancer incidence in smokers, ex-smokers, and nonsmokers, it has been posited that cigarette smoke induces both the initial and penultimate changes of a five- or six-stage Armitage-Doll model. In this paper we fit the Armitage-Doll model under the assumption that both the initial and penultimate changes are induced by cigarette smoke to the data of Doll and Hill and show that this dual role for cigarette smoke does not explain the incidence curves in smokers and ex-smokers when considered jointly. In addition, a fundamental dose-response implication of this dual-role model is derived and shown not to be in accord with the data. Furthermore, we show in a theoretical manner why this model is not flexible enough to explain the diverse patterns of incidence and, hence, is not an adequate model for lung carcinogenesis.

In light of the problems of the Armitage-Doll model, we propose a simple two-stage model with clonal growth in which cigarette smoke induces the initial transition and promotes clonal growth in those cells initiated by cigarette smoke. This model fits the incidence of lung cancer in both smokers and ex-smokers and its predicted dose-response also conforms with the data. The model is proposed on this em-

[1]Received November 12, 1987; revised May 18, 1988; accepted May 20, 1988.

[2]Supported in part by Public Health Service grants ES-00260 from the National Institute of Environmental Health Sciences and CA-13343 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services; and by grant SIG-00009 from the American Cancer Society.

[3]Department of Clinical and Scientific Affairs, Pfizer Inc., 235 East 42nd St., New York, NY 10017.

[4]Institute of Environmental Medicine, New York University Medical Center, New York, NY.

[5]The authors thank Dr. Charles Brown and Dr. Alice Whittemore for their helpful comments.

pirical basis. Multistage models incorporating clonal growth were first proposed by Armitage and Doll (*4*) and Fisher (*5*). More recently, Moolgavkar and Knudsen (*6*) proposed a two-stage model with clonal growth that they state is sufficient to generate a wide range of background incidence curves.

The remainder of this introductory section gives some background on the hypothesis that the two affected stages of an Armitage-Doll model explain the incidence rate of lung cancer in smokers and ex-smokers. In the next section, we show that the model does not fit the observed incidence rates in smokers and ex-smokers. The third section deals with the dose-response implications of the model. Then in the fourth section we propose the clonal-growth model and show the agreement between this model and the data. We end with a discussion of the results.

In his initial analysis of the first 17 years of observation in the Doll and Hill study, Doll (*1*) found (*a*) that the excess incidence rate of lung cancer in smokers is proportional to a high power of smoking duration in agreement with the relationship between lung cancer incidence and age in nonsmokers; (*b*) that the excess incidence rate of lung cancer in ex-smokers seems to remain constant after smoking stops; and (*c*) that the results suggest a linear relationship between incidence and amount smoked.

In the discussion that followed the presentation of these results by Doll to the Royal Statistical Society in 1970, Armitage (*7*) discussed these results in the context of the Armitage-Doll model. He noted that the incidence rate while smoking continued indicated that the initial transition rate was increased. He also pointed out that if only the first stage were involved then the rate after smoking ceases would continue to rise in the same way as if smoking continued, whereas the observed stabilization of incidence rate is what would be expected if the next-to-last stage were affected. He expressed surprise, however, that the relationship between amount smoked and incidence appeared linear, which suggested that one stage was affected rather than two.

The same data were analyzed in detail by Whittemore and Altshuler (*8*), who found incidence rate to be proportional to between the fourth and fifth power (4.7) of smoking duration and linear in amount smoked. Whittemore (*9*) pointed out that the power relationship between the bronchial carcinoma incidence rate of smokers and smoking duration is the same as that between nonsmokers and age.

Doll and Peto (*10*) analyzed a subset of 20 years of observations from the Doll and Hill study. They also found lung cancer incidence rate to be proportional to between the fourth and fifth power (4.5) of smoking duration. It is stated in the paper that if cancers arise from cells that were normal until acted on by cigarette smoke, the Armitage-Doll model predicts that incidence rate should rise approximately proportional to some power of the duration of smoking. They also state that the effect of giving up smoking suggests that smoking also affects a late stage in the process. More important, in contrast to Doll's (*1*) initial finding of a linear dose-response relationship and in contrast to the analysis of Whittemore and Altshuler (*8*), Doll and Peto suggest a pure quadratic dose-response relationship between cigarette smoke and lung cancer incidence. Further, they state that this is what might be expected if more than one of the stages was strongly affected by smoking. Thus, they suggest that the anomaly raised by Armitage (*7*), that the incidence in smokers and ex-smokers indicates that cigarette smoke affects two stages but the dose-response relationship indicates only one, is removed as evidence counter to the hypothesis that cigarette smoke affects two stages. A similar view of these data is reached by Day and Brown (*11*) in a paper reviewing incidence data in the context of multistage models. Whittemore (*9*) refers to the same evidence, which supports both the initial and penultimate rates.

In a comprehensive review of the multistage model, Peto (*12*) examined lung cancer death rates in British females who began smoking at varying ages, and he concluded that one of the necessary cellular changes caused by smoking occurs many years before the appearance of lung cancer (consistent with an effect on an early cellular change) while the effect of giving up smoking seems to implicate the penultimate transition rate. In addition, Peto clearly enumerates the many biases that may lead to underestimating the exponent of dose (number of cigarettes smoked per day) and concludes that the data are more consistent with a quadratic relationship than a linear one.

In two recent papers, Freedman and Navidi (*13,14*) examined the fit of the multistage model to three lung cancer data sets including the Doll and Hill data. They conclude that the multistage model may not describe the carcinogenic process. Furthermore, using data from the two other data sets, they conclude that excess incidence in fact declines in ex-smokers and state this is incompatible with the Armitage-Doll model.

Results

Incidence Rate in Smokers and Ex-Smokers

Under the multistage model of Armitage and Doll, the background incidence rate of lung cancer at age x is given by:

$$N\lambda_1 \ldots \lambda_k x^{k-1}/(k-1)! \qquad [1]$$

where $\lambda_i > 0$ is the background transition rate of a cell in stage i-1 going to stage i, N is the number of normal cells in the tissue at zero age, and k is the number of cellular changes. (The constant $N\lambda_1 \ldots \lambda_k/(k-1)!$ will be denoted by c.) Background incidence is the result of the amalgam of causes, either biological or environmental (other than cigarette smoke), which yield constant transition rates for the k cellular changes.

Assuming that cigarette smoke induces the first and penultimate changes, then for a fixed dose (number of cigarettes per day) the first transition rate and the penultimate transition rate can be expressed as $(1 + \gamma_1)\lambda_1$ and $(1 + \gamma_{k-1})\lambda_{k-1}$ while the other transition rates remain at background levels. Under these conditions it can be shown (*15,16*) that the excess incidence rate of lung cancer in continuing smokers of age x is:

$$c[\gamma_1(x-x_0)^{k-1} + \gamma_{k-1}(x^{k-1} - x_0^{k-1}) + \gamma_1\gamma_{k-1}(x-x_0)^{k-1}] \qquad [2]$$

Table 1. No. of lung cancers and person-years by age and amount smoked for continuing smokers*

Age (yr)	Cigarettes/day (mean)†						
	Never smoked	10–14 (11.3)	15–19 (16.0)	20–24 (20.4)	25–29 (25.4)	30–34 (30.2)	35–40 (38.0)
40–44	0/17,846.5	1/3,795.5	0/4,824	1/7,046	0/2,523	1/1,715.5	0/892.5
45–49	0/15,837.5	1/3,205	1/3,995	1/6,460.5	2/2,565.5	2/2,123	0/1,150
50–54	1/12,226	2/2,727	4/3,278.5	6/5,583	3/2,620	3/2,226.5	3/1,281
55–59	2/8,905.5	1/2,288	0/2,466.5	8/4,357.5	5/2,108.5	6/1,923	4/1,063
60–64	0/6,248	1/1,714	2/1,829.5	13/2,863.5	4/1,508.5	11/1,362	7/826
65–69	0/4,351	2/1,214	2/1,237	12/1,930	5/974.5	9/763.5	9/515
70–74	1/2,723.5	4/862	4/683.5	10/1,055	7/527	2/317.5	5/233
75–79	2/1,772	4/547	5/370.5	7/512	4/209.5	2/130	2/88.5

*Taken from tables 2 and 3 of Doll and Peto (*10*).
† Values = No. of lung cancers/No. of person-yr.

where x_0 is the age smoking began and in agreement with Doll and Peto (*10*) is taken to be 22.5 for the Doll and Hill data.

The first term in expression 2 gives the excess incidence if smoking induced only the initial transition; the second term gives the excess incidence if smoking induced only the penultimate transition; the third term gives the excess incidence rate due to the joint action of cigarette smoke at the initial and penultimate stages and will be referred to as the multiplicative component. One sees from expression 2 that while smoking continues, the excess incidence due to the initial transition component and the multiplicative component rise to the k−1 power of smoking duration while the penultimate transition component rises less rapidly than the k−1 power of duration. Because of this and because the power of smoking duration has been found to be between 4.5 and 5 (*8,10*) in the Doll and Hill data, expression 2 with k = 5 cannot fit the data as well as k = 6. Thus, we determine the best fit of expression 2 to the data for a six-stage process.

Table 1 contains the basic incidence data given in Doll and Peto (*10*) for six dose and eight age categories and for life-long nonsmokers. In order to estimate the parameters in expression 2, for each age group the data were combined by the same weighted average of the six dose-specific incidences. The weight given each dose group is the number of person-years in that dose group divided by the total number of person-years over all doses. Hence, these combined age-specific incidences are standardized for dose, and applying these weights to the mean number of cigarettes given in each dose group yields 20.8 cigarettes per day. Thus, we take the combined incidences given in table 2 to be the incidence of lung cancer in smokers of 20.8 cigarettes per day at varying ages with smoking starting at age 22.5.

Since γ_1 and γ_5 of expression 2 represent proportional increases over the transition rates in nonsmokers, the incidence rates in nonsmokers are used to obtain a proper constraint for γ_1 and γ_5. By summing the incidences given in table 1 for nonsmokers and those given in table 2 for smokers, we estimate the cumulative excess incidence from age 40 to age 80 in continuing smokers to be 18.7 relative to that of life-long nonsmokers over the same age range. Integrating expressions 1 and 2 from ages 40 to 80 and dividing the results of the two integrations lead to the following constraint imposed by the model:

$$.140(\gamma_1 + \gamma_1 \gamma_5) + .995 \gamma_5 = 18.7 \quad [3]$$

Solving the above expression for γ_1 and inserting the result in expression 2 yields:

$$c[((18.7 - .995 \gamma_5)/.140)(x-x_0)^5 + \gamma_5 (x^5 - x_0^5)] \quad [4]$$

for the excess incidence rate, and this expression 4 was fit to the data in table 2. Strictly speaking, the incidences in table 2 are not excess incidences since no adjustment for background incidence has been made. However, any adjustment would be negligible and we did not adjust in this data set because of the small numbers of cancers in any 5-year age interval for nonsmokers. The Gauss-Newton method was used to obtain the estimates of c and γ_5, which minimized the sum of squared deviations between the observed and predicted values. γ_5 is then inserted into the constraint to obtain γ_1. As shown in table 3, this best fit of the six-stage Armitage-Doll model gives an accurate description of the observed incidence in continuing smokers.

It remains to be examined if this fit also predicts the relatively constant excess incidence rate seen in ex-smokers. The Armitage-Doll model predicts excess incidence rate in ex-smokers of age x to be:

$$c[\gamma_1((x-x_0)^5-(x-x_1)^5) + \gamma_5(x_1^5 - x_0^5) + \gamma_1\gamma_5(x_1 - x_0)^5] \quad [5]$$

where x_1 is the age smoking stopped (*11,15*). Analogous to the description given after expression 2 for continuing smokers, the three terms given in expression 5 are the initial stage component, the penultimate stage component, and the multiplicative component. As seen in expression 5, the penultimate component and the multiplicative component depend on smoking duration and remain constant after smoking

Table 2. Dose-standardized incidence (10^5 person-yr) by age in continuing smokers*

	Age (yr)							
	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79
Incidence	15.6	36.5	116.6	157.2	356.0	571.0	886.1	1,404.7

*Dose-standardized incidence was computed by a weighted average of the 6 dose-specific incidences given in table 1. The weight given each dose-group is the No. of person-yr in that dose group (over all age groups) divided by the total No. of person-yr.

Table 3. Observed incidence (10^5 person-yr) in continuing smokers and predicted incidence of 3 fits of the Armitage-Doll model with cigarette smoke inducing the first and penultimate changes

Age (yr)	Observed incidence (from table 2)	Best 6-stage fit $c = 2.23 \times 10^{-8}$ $\gamma_1 = 14.3, \gamma_5 = 5.6$	Maximum multiplicative 6-stage fit $c = 2.26 \times 10^{-8}$ $\gamma_1 = 10.4, \gamma_5 = 7.0$	Best 7-stage fit $c = 3.25 \times 10^{-10}$ $\gamma_1 = 3.1, \gamma_5 = 14.1$
42.5	15.6	23.2	27.1	27.3
47.2	36.5	49.7	55.9	55.5
52.5	116.6	99.7	108.2	105.9
57.5	157.2	187.3	197.7	192.0
62.5	356.0	332.1	343.2	330.0
67.5	571.0	559.8	568.9	555.8
72.5	886.1	902.6	905.2	897.0
77.5	1,404.7	1,400.8	1,390.2	1,404.3
Residual sum of squares	—	2,408	2,968	2,697

stops, while the initial stage component continues to increase. Because of this, the Armitage-Doll model which gives the best fit for continuing smokers does not predict that excess incidence rate remains constant after smoking stops, but predicts a continued rise with the excess incidence (relative to that of stopping age) depending little on the age when smoking stops. For example, table 4 shows the rise in excess incidence in ex-smokers who stopped at age 52.5. Twenty years after smoking stopped the excess incidence would be almost double that of the age when smoking stopped.

We examined the data to see if excess incidence rate does in fact remain constant after smoking stops. Table 5 contains the observed number of lung cancers in ex-smokers taken from table 13 of Doll (*1*). In addition, we calculated the expected number of cancers in ex-smokers based on the observed incidence in continuing smokers at the age when smoking stopped. The cumulative results are consistent with a constant excess incidence when smoking stops (or a small decline) but give no indication of the large continued rise in excess incidence as predicted by the Armitage-Doll model, which best fits incidence in continuing smokers.

Although the best fit of the Armitage-Doll model in continuing smokers does not fit the data in ex-smokers, the fit in continuing smokers can be loosened to a small degree in order to improve the fit in ex-smokers and thereby obtain a better simultaneous fit. This can be accomplished both within the six-stage model and by increasing the number of stages.

Table 4. Predicted incidence in ex-smokers who stopped smoking at age 52.5 for 3 fits of the Armitage-Doll model*

Yr after smoking stopped	Fit		
	1	2	3
5	1.09	1.06	1.01
10	1.25	1.17	1.03
15	1.51	1.35	1.07
20	1.92	1.62	1.14
25	2.51	2.01	1.25

*Incidence is expressed relative to the incidence at age 52.5. Fit 1 is the best 6-stage fit; fit 2 is the maximum multiplicative 6-stage fit; and fit 3 is the best 7-stage fit.

Within the six-stage model, the values of γ_1 and γ_5 that maximize the multiplicative component of excess incidence will give the best simultaneous fit. This is so because only the multiplicative component agrees with the time pattern of lung cancer incidence in both smokers and ex-smokers. Thus, we maximized the quantity $\gamma_1 \gamma_5/(\gamma_1 + \gamma_5 + \gamma_1 \gamma_5)$ subject to the constraint given in expression 3. As seen in tables 3 and 4, while the fit is somewhat worsened in continuing smokers, the rise of excess incidence in ex-smokers is attenuated. Nevertheless, the predictions of the model still are not in accordance with the data for ex-smokers. Thus, the six-stage Armitage-Doll model is simply not flexible enough to fit rising incidence in smokers and constant incidence in ex-smokers simultaneously.

The best fit of a seven-stage model to the incidence in continuing smokers was also obtained. As shown in table 3, this model fits the data almost as well as the best six-stage fit in continuing smokers and, as seen in table 4, predicts excess incidence in ex-smokers which rises less rapidly than that seen for the six-stage fit. Hence, a seven-stage model gives a better simultaneous fit in smokers and ex-smokers, and, it is important to note, the estimated values of γ_1 and γ_6 indicate that in this model cigarette smoke induces lung cancer predominantly by acting on the penultimate stage. In fact, for an eight-stage model the penultimate stage component alone might be considered to fit the data for continuing smokers (residual sum of squares is 3,943) and, as stated above, this component remains constant when smoking stops. Thus, paradoxically within the Armitage-Doll model, a single role for cigarette smoke, albeit with a higher number of stages, explains the joint data better than does the dual role that was hypothesized for the express purpose of explaining the diverse patterns of incidence in smokers and ex-smokers. However, the higher number of stages is to be ruled out because it implies too high a power dependence of incidence with age in nonsmokers.

Dose-Response

As discussed previously, the Doll and Hill data have been interpreted as showing either a linear or quadratic relationship between dose (number of cigarettes smoked per day)

Table 5. Observed No. of lung cancers in ex-smokers and expected No. given constant excess incidence after smoking stops

Age (yr) stopped smoking	Amount smoked (cigarettes/day)							
	25+		15–24		1–14		Total	
	Observed*	Expected†	Observed	Expected	Observed	Expected	Observed	Expected
45–54	6	4.9	1	3.4	2	1.6	9	9.9
55–64	10	7.9	4	5.3	2	2.1	16	15.3
65–74	5	6.1	1	5.6	3	3.9	9	15.6
Total	21	18.9	6	14.3	7	7.6	34	40.8

*Taken from table 13 of Doll (*1*).

†The expected No. was calculated in the following way: the No. of lung cancers for the appropriate age and dose groups were summed and divided by the corresponding No. of patient-yr given in tables 2 and 3 of Doll and Peto (*10*) for continuing smokers. This incidence was then multiplied by the appropriate No. of patient-yr for ex-smokers given in table 13 of Doll (*1*).

and lung cancer incidence (*8,10*). In this section we examine the dose-response relationship under the Armitage-Doll model with the relative increases in the initial and penultimate transition rates proportional to dose. The excess incidence rate in continuing smokers given in expression 2 can be rewritten as:

$$c[\alpha_1 d(x-x_0)^{k-1} + \alpha_{k-1}\, d(x^{k-1}-x_0^{k-1}) + \alpha_1\alpha_{k-1}\, d^2\,(x-x_0)^{k-1})] \quad [6]$$

where d denotes number of cigarettes per day.

One sees that the model predicts a dose-response relationship which is a second degree polynomial in dose. Only the multiplicative component varies as d^2 while the initial and penultimate components are linear in dose. Because of the different time patterns of these components the coefficient of the d^2 term, $\alpha_1\alpha_{k-1}\,(x-x_0)^{k-1}$, increases relative to the coefficient of the linear term as smoking duration increases. Hence the dose-response relationship is not constant but depends on smoking duration and becomes more nearly quadratic as smoking duration increases. Because of this dependence of dose-response on duration, the best fit (six-stage) of the Armitage-Doll model in continuing smokers given in the previous section predicts that for smokers aged 42.5 (smoking for 20 years) the effect of smoking twice as much (approximately two packs per day) is to increase the incidence by a factor of 2.5. However, in smokers aged 72.5 (smoking for 50 years) the effect of smoking twice as much is to increase the excess incidence rate by 3.3.

This prediction is not supported by the data. Whittemore and Altshuler (*8*) examined the dose-response relationship for four dose categories within each of five duration-of-smoking categories from the Doll and Hill data. They showed that the relationship of incidence proportional to the 1.1 power of amount smoked holds for each of the five duration-of-smoking categories indicating that the dose-response relationship does not depend on smoking duration. We examined the data in a simpler way. Table 6 contains the incidence rates for two dose categories and four duration categories. These incidences were obtained by a weighted average of the incidences within the three lower dose groups and the three higher dose groups given in table 1, where the weights are proportional to the number of person years. These results show that the proportional increase in incidence from low to high dose is relatively constant from 17.5 to 47.5 years of smoking and is reduced in the longest duration of smoking category, contrary to the prediction of the model. Thus, the predictions of the best fit in continuing smokers of the six-stage Armitage-Doll model with cigarette smoke inducing the initial and penultimate changes are contradicted by dose-response results as well as the incidence in ex-smokers.

For the best simultaneous fit in smokers and ex-smokers, the eight-stage fit of the penultimate component only, the dose-response is independent of duration and is linear. However, table 6 shows an increase in incidence that is more than proportional to the increase in mean number of cigarettes per day in three of the duration groups. Thus, nonlinearity of dose-response actually runs counter to this possible fit of the Armitage-Doll model.

Two-Stage Model with Clonal Growth

In this section we show that a two-stage model for cigarette-induced cancer with clonal growth after the initial change fits the data in continuing smokers as well as the Armitage-Doll model and predicts a constant excess incidence after smoking stops. A similar model was first proposed by Fisher (*5*).

Assume that two cellular changes are required in lung carcinogenesis and that cigarette smoke induces the initial transition and promotes clonal growth in those cells induced by cigarette smoke so that the number of cells in the clone can be approximated by some power b of the age of the clone. Under these assumptions it is shown (*15*) that the excess incidence of lung cancer in continuing smokers of age x is:

$$a(x-x_0)^{b+1} \quad [7]$$

Table 6. Lung cancer incidence (10^5 person-yr) for 2 dose and 4 smoking duration categories

Smoking duration (yr)	Cigarettes/day (mean)		Ratio
	10–24 (16.8)	25–40 (29.7)	
17.5–27.5	17.6	44.3	2.5
27.5–37.5	99.2	219.2	2.2
37.5–47.5	316.3	789.7	2.5
47.5–57.5	961.5	1,547.5	1.6

where a is a constant that depends on the two transition rates and the rate of clonal growth.

Expression 7 was fit to the data in table 2 by the Gauss-Newton method and the results are shown in table 7. As indicated by a comparison of the residual sum of squares of this fit with those in table 3, the clonal-growth model fits the data in continuing smokers as well as does the Armitage-Doll model. Furthermore, when smoking stops, the model predicts that excess incidence rate will remain constant. This is so because the initial transition is also the penultimate one in a two-stage model and the growth of the clone stops when smoking stops. Thus, this model fits the diverse patterns of incidence in smokers and ex-smokers.

The dose-response relationship predicted by the clonal-growth model is determined jointly by the relationship of dose to the induction of the initial change and to the rate of growth of the clone. If each relationship is linear then the dose-response is quadratic; however, if only the initiation effect is linear and the proliferative effect is less than linear, then the dose-response relationship is between linear and quadratic. More important, this dose-response relationship does not depend on smoking duration, unlike the dose-response in the Armitage-Doll model and in agreement with the data shown in table 6.

Thus, this clonal-growth model is in accordance with the incidence data for smokers and ex-smokers and for dose-response. Furthermore, as suggested by the combined data in table 5, excess incidence may even decrease in ex-smokers. This was the finding of Freedman and Navidi (*14*) for lung cancer in ex-smokers. Our model is easily modified to be consistent with this finding by allowing the size of the clone to shrink after smoking stops.

The critical assumptions in our model are that the cigarette smoke does not act on cells initiated by other agents and that the growth of the cigarette-induced clones stops when smoking stops. Without these assumptions there is an equivalence between the Armitage-Doll model and the clonal-growth model (*15*). Specifically, if cigarette smoke increased both the initial transition rate and a background growth rate of all clones then the model would have predictions of excess incidence in smokers and ex-smokers identical to those of a b + 2-stage Armitage-Doll model in which cigarette smoke increased both the initial and penultimate rates.

Table 7. Observed incidence (10^5 person-yr) in continuing smokers and predicted incidence of clonal-growth model

Age (yr)	Observed incidence	Best clonal-growth fit $a = 2.13 \times 10^{-5}$, $b = 3.49$
42.5	15.6	14.8
47.5	36.5	40.4
52.5	116.6	91.7
57.5	157.2	183.2
62.5	356.0	333.7
67.5	571.0	566.4
72.5	886.1	909.2
77.5	1,404.7	1,394.9
Residual sum of squares		2,458

Discussion

The results presented in this paper show that the Doll and Hill lung cancer data are not consistent with the widely held hypothesis that cigarette smoke induces the first and penultimate stages of a six-stage or less Armitage-Doll model of carcinogenesis. This model can not simultaneously fit incidence in smokers and ex-smokers. The best fit in smokers predicts that excess incidence will greatly increase in ex-smokers, whereas the data indicate no change or a decrease. In addition, the model predicts greater nonlinearity in the dose-response relationship with greater smoking duration, whereas the data do not show this. A better simultaneous fit in smokers and ex-smokers can be obtained by a higher number of stages, but this de-emphasizes initiation and is inconsistent with the dependence of incidence on age in nonsmokers.

That the data do not support the Armitage-Doll model is the paramount result of this paper. Because of this result, we propose a clonal-growth model in which cigarette smoke initiates and promotes clonal growth in those cells initiated by cigarette smoke, but does not act on any background-initiated cells. This model fits the incidence data in smokers and ex-smokers and is proposed on this empirical basis. Clonal growth, per se, is well supported in the literature and generally accepted to play a role in carcinogenesis. Our finding that the data are consistent with cigarette smoke not acting on background-initiated cells, while controversial, should motivate the study of clonal growth, although the reasons for this apparent independence may be temporal (time pattern of background exposure) or spatial (location in the lung) as well as biological. The strength of the model lies in its being the simplest theoretical statement that accords with the data.

In his response to the conflicting evidence of the role of cigarette smoke in lung cancer raised by Armitage, Doll (*1*) called this anomaly crucial and indicated that its resolution would result in a much clearer understanding of the mechanism of cancer induction. The results of this paper show that the Armitage-Doll model is not an adequate description of the mechanism. Moreover, a high number of stages has not been experimentally demonstrated. In contrast, a two-stage model with clonal growth is supported by experimental data. Two examples are the hyperplastic foci reported by Shimkin and Polissar (*17*) in their investigation of pulmonary carcinogenesis and the data of Albert et al. (*18*), which indicate that papillomas progress to carcinomas. Moolgavkar and Knudson (*6*) concluded that the most important effect of cigarette smoke is its proliferative effect, and they point out that there is evidence that cigarette smoke causes hyperplasia of the bronchial epithelium (*19*) and that tobacco extract is known to contain tumor promoters (*20*).

Other carcinogens have been examined in the context of the multistage model. Day and Brown (*11*) review both experimental and epidemiological data and conclude that both early and late-stage effects were seen. For example, the role of nickel in inducing nasal sinus tumors appears to be a late-stage one. Brown and Chu (*21*) used multistage theory to study occupational exposure to arsenic, and they conclude

that this carcinogen appears to act at a late stage. Collingwood et al. (22) found that the time pattern of respiratory cancer in chemical workers suggests that chloromethyl ethers affect both early and late stages. Since the time pattern of carcinogens that act by promoting clonal growth of initiated cells is the same as that of carcinogens that act on the penultimate stage of a multistage process, nickel, arsenic, and chloromethyl ethers may act by promoting clonal growth, rather than by affecting a late stage of a multistage process.

The difference between these two models of carcinogenesis is important from the perspective of prevention, since, if clonal growth does contribute to the high exponent seen in cancer incidence curves, then it plays a dominant role in the process; that is, the chance of at least one normal cell proceeding to complete transformation without clonal growth would be relatively small. In other words, the transition rates alone are so small as to make the appearance of a tumor within life span very improbable. Therefore, a method that would eliminate or decrease the selective advantage of partially transformed cells would be most effective in preventing tumors. Thus, in contrast to the usual sequence where experimental results suggest epidemiology studies, in this case the epidemiology of lung cancer should motivate the experimentalist to study clonal growth.

References

1. DOLL R. The age distribution of cancer: implications for models of carcinogenesis. J R Stat Soc Ser A 1971;134:133–155.
2. DOLL R, PETO R. Mortality in relation to smoking: 20 years' observations on male British doctors, Br Med J 1976;2:1525–1536.
3. ARMITAGE P, DOLL R. The age distribution of cancer and a multi-stage theory of carcinogensis, Br J Cancer 1954;8:1–12.
4. ARMITAGE P, DOLL R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. Br J Cancer 1957;11:161–169.
5. FISHER JC. Multiple-mutation theory of carcinogenesis. Nature 1958;181:651–652.
6. MOOLGAVKAR SH, KNUDSON AG, Jr. Mutation and cancer: a model for human carcinogenesis. JNCI 1981;66:1037–1052.
7. ARMITAGE P. Discussion on Doll's paper. J R Stat Soc Ser A 1971;134:155–156.
8. WHITTEMORE A, ALTSHULER B. Lung cancer incidence in cigarette smokers: further analysis of Doll and Hill's data for British physicians. Biometrics 1976;32:805–816.
9. WHITTEMORE A. Epidemiologic implications of the multistage theory of carcinogenesis. In: Environmental health quantitative methods. Philadelphia: Society for Industrial and Applied Mathematics, 1977:72–87.
10. DOLL R, PETO R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. J Epidemiol Community Health 1978;32:303–313.
11. DAY NE, BROWN CC. Multistage models and primary prevention of cancer. JNCI 1980;64:977–989.
12. PETO R. Epidemiology, multistage models and short-term mutagenicity tests. In: Origins of human cancer. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1977:1403–1428.
13. FREEDMAN DA, NAVIDI W. On the multistage model for carcinogenesis. Technical Report No. 47, University of California (1987).
14. FREEDMAN DA, NAVIDI W. On the risk of lung cancer for ex-smokers. Technical Report No. 135, University of California (1987).
15. GAFFNEY M. Quantitative implications of the multistage model for environmentally induced cancer. PhD thesis. New York: New York University, 1983.
16. WHITTEMORE A. The age distribution of human cancer for carcinogenic exposure of varying intensity. Am J Epidemiol 1977;106:418–432.
17. SHIMKIN MB, POLISSAR MJ. Some quantitative observations on the induction and growth of primary pulmonary tumors in strain A mice receiving urethane. J Natl Cancer Inst 1955;16:75–94.
18. ALBERT RE, BURNS FJ, ALTSHULER B. Reinterpretation of the linear nonthreshold dose-response model in terms of the initiation-promotion mouse skin tumorigenesis. In: New concepts in safety evaluation, part II. New York: Halsted Press, 1979:89–95.
19. AUERBACH O, STOUT AP, HAMMOND EC, et al. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. N Engl J Med 1961;265:253–267.
20. VAN DUUREN BL, GOLDSCHMIDT BM. Cocarcinogenic and tumor-promoting agents in tobacco carcinogenesis. J Natl Cancer Inst 1976;56:1237–1242.
21. BROWN CC, CHU KC. Implications of the multistage theory of carcinogenesis applied to occupational arsenic exposure. JNCI 1983;70:455–463.
22. COLLINGWOOD KW, PASTERNACK BS, SHORE RE. An industry-wide study of respiratory cancer in chemical workers exposed to chloromethyl ethers. JNCI 1987;78:1127–1136.